Confronting weather and climate models with observational data from soil moisture networks over the United States

Paul A. Dirmeyer[1], Jiexia Wu[1], Holly E. Norton[1], Wouter A. Dorigo[2,3], Steven M. Quiring[4], Trenton W. Ford[5], Joseph A. Santanello Jr.[6], Michael G. Bosilovich[6], Michael B. Ek[7], Randal D. Koster[6], Gianpaolo Balsamo[8], and David M. Lawrence[9]

[1]George Mason University, Fairfax, VA, USA

[2]Vienna University of Technology, Vienna, Austria

[3]Laboratory of Forest and Water Management, Ghent University, Ghent, Belgium

[4]Texas A&M University, College Station, TX, USA

[5]Southern Illinois University, Carbondale, IL, USA

[6]NASA Goddard Space Flight Center, Greenbelt, MD, USA

[7]NOAA National Centers for Environmental Prediction, College Park, MD, USA

[8]European Centre for Medium-range Weather Forecasts, Shinfield Park, Reading, UK

[9]National Center for Atmospheric Research, Boulder, CO, USA

Corresponding Author:

Paul A. Dirmeyer
Center for Ocean-Land-Atmosphere Studies
George Mason University
4400 University Drive, Mail Stop: 6C5
Fairfax, Virginia 22030 USA
pdirmeye@gmu.edu

Revised: 13 January 2016
Submitted to: *Journal of Hydrometeorology*

Abstract

Four land surface models in uncoupled and coupled configurations are compared to observations of daily soil moisture from 19 networks in the conterminous United States to determine the viability of such comparisons and explore the characteristics of model and observational data. First, observations are analyzed for error characteristics and representation of spatial and temporal variability. Some networks have multiple stations within an area comparable to model grid boxes; for those we find that aggregation of stations before calculation of statistics has little effect on estimates of variance, but soil moisture memory is sensitive to aggregation. Statistics for some networks stand out as unlike those of their neighbors, likely due to differences in instrumentation, calibration and maintenance. Buried sensors appear to have less random error than near-field remote sensing techniques, and heat dissipation sensors show less temporal variability than other types.

Model soil moistures are evaluated using three metrics: standard deviation in time, temporal correlation (memory) and spatial correlation (length scale). Models do relatively well in capturing large-scale variability of metrics across climate regimes, but poorly reproduce observed patterns at scales of hundreds of kilometers and smaller. Uncoupled land models do no better than coupled model configurations, nor do reanalyses outperform free-running models. Spatial decorrelation scales are found to be difficult to diagnose. Using data for model validation, calibration or data assimilation from multiple soil moisture networks with different types of sensors and measurement techniques requires great caution. Data from models and observations should be put on the same spatial and temporal scales before comparison.

1. Introduction

Coupled land-atmosphere model development has lagged behind coupled ocean-atmosphere model development for a variety of reasons. Top among them is that the necessary measurements for assessing land-atmosphere feedback processes have been largely lacking. In recent years, co-located measurements of surface fluxes, near surface meteorology and land surface states like soil moisture have begun to cross a critical threshold of quantity and coverage, largely due to the maturation of the global FluxNET set of environmental measurements (Baldocchi et al. 2001). Systematic benchmarking of land surface models (LSMs) has begun based on simulation of daily mean surface fluxes (Best et al. 2015). However, for soil moisture alone there are even more widespread data in the form of many independent networks of *in situ* measurements (Dorigo et al. 2011, Quiring et al. (under review)). They span a tremendous range of station densities, down to sub-grid scales relative to current weather and climate models, making them enticing for model calibration and validation.

Bringing observational data to bear on model improvement requires not just the data sets and models themselves, but also sound methods of analysis and processes understanding to guide the approach. Comparing models with observations can easily become misguided, if not actually unfair, if the basic differences between how models represent the world and how instruments measure the world are not carefully considered and accounted for. For a quantity like soil moisture, this is a particularly significant issue (Dirmeyer 2004, Koster et al. 2009). Xia et al. (2015) spatially averaged both model and observed data to coarse scales to facilitate comparison. Stillman et al. (2014) assessed the ability of multiple soil moisture instruments in a catchment to represent area-averaged soil moisture, using a higher

density raingauge network to infer smaller scale variations. Gruber et al. (2013) examined random errors in soil moisture at spatial scales comparable to global models using a triple colocation method combining remotely sensed and modeled soil moisture estimates with *in situ* soil moisture measurements, highlighting the care that must be taken in applying *in situ* measurements as ground truth. Such approaches hold promise to evaluate remote sensing products (Dorigo et al. 2015) and improve estimates of soil moisture–atmosphere interactions (Crow et al. 2015).

In this study we confront 12 unique model configurations using four different land surface models with soil moisture measurements from 19 networks across the conterminous United States. However, we first address the observational data sets themselves to estimate their error characteristics in a distinctive way based on lagged autocorrelation statistics, and their representativeness of temporal variability, spatial and temporal scales.

Section 2 describes the observational and model data used. The metrics evaluated are introduced in Section 3, and Section 4 presents an evaluation of observational error. Scaling issues are addressed in Section 5. Section 6 gives an evaluation of soil moisture variance and memory in observations and models. Spatial scales of soil moisture variability are considered in Section 7, and conclusions and summary are offered in Section 8.

2. Data

In this comparison, point observations and model grid-box estimates of soil moisture data at daily time intervals or daily time means are used. The domain of observations for this study is confined to the conterminous United States, and model

75 comparisons are performed over roughly the same area. Table 1 lists all networks
76 used, the data collections (described below) from which data were taken, the
77 location of the networks (many are regional) and the type of instrumentation each
78 uses.

79 *2.1 International Soil Moisture Network*

80 The International Soil Moisture Network (ISMN) is a data synthesis effort focused on
81 collecting in situ soil moisture measurements and associated co-located
82 observations of relevant meteorological data from all available international sources
83 (Dorigo et al. 2011; 2013). The resulting quality-controlled database of raw
84 observations is meant to provide ground-truth calibration and validation for
85 satellite observations as well as for the calibration and validation of land surface
86 models. It is coordinated by the Global Energy and Water Exchanges Project
87 (GEWEX) in cooperation with the Group of Earth Observation (GEO) and the
88 Committee on Earth Observation Satellites (CEOS).

89 Data from many different networks and extended field campaigns are archived by
90 ISMN. Those networks used in this experiment are listed in Table 2. ISMN archives
91 data at the highest available temporal resolution up to hourly from each reporting
92 instrument, allotting one file for each instrument and level. Basic quality control is
93 performed and records suspected to be out-of-range or otherwise untrustworthy
94 are flagged, but no data is omitted. Fig 1a shows the locations of stations used from
95 ISMN.

96 *2.2 North American Soil Moisture Database*

97 The North American Soil Moisture Database (NASMD; Quiring et al. (2015)) is a
98 collection of harmonized daily soil moisture data from in situ measurements across

North America. Networks used in this study from NASMD are listed in Table 3. The motivation for NASMD is to provide a data set to investigate processes by which soil moisture variability influences climate on seasonal to interannual timescales over North America. Unlike ISMN, NASMD provides processed data from each station location in each network. Daily values are calculated from stations with sub-daily data using a simple average. Interpolation is used to fill gaps of less than 10 days using a monthly average replacement method, which has been shown to work well with daily observations (Ford and Quiring 2014).

Only one time series is provided at each station and sensor depth, regardless of how many instruments are in place. When stations have multiple instruments at a single depth, usually the first reported sensor is used. However, if data from the first instrument are flagged by the quality control routine or if there are excessive missing observations, the next reported instrument is considered and may be used instead. Because of micro-scale variability in soil texture, averaging observations from multiple sensors was considered unjustifiable. Fig 1b shows the distribution of stations in the NASMD repository.

2.3 Observational Data Processing

The data from both collections were further processed for this study. Each data file is scanned for basic statistics including the time range of available data, depths of instrument readings, and data reporting intervals so the data from each network can be synthesized into a single file spanning the maximum time range of the network's observations. For the ISMN data, daily means are first calculated.

For each station and profile of sensors in the soil (or across the reporting depth for remote sensors like in COSMOS or PBO-H2O), the observational data are vertically interpolated to the model levels for each of the four land surface schemes in this

study, following the procedure used in the second Global Soil Wetness Project (Dirmeyer et al. 2006). The lowest model layer to encompass the depth of the deepest reporting sensor is the lowest model layer to contain interpolated data – layers below are set to missing. Model layers above the shallowest sensor depth are set to the soil moisture value of that shallowest sensor. Otherwise, it is assumed that the observed data of buried sensors are representative of a layer whose top is exactly halfway between it and the next shallowest sensor (or the surface if it is the shallowest sensor), and whose bottom is exactly halfway between it and the next deepest sensor (or if it is the deepest sensor, to the same distance below as the top boundary was determined to be above it). It is assumed that the soil moisture across this thickness is uniform, as is typically supposed for land surface model layers. Then the interpolated value for any model layer is a simple weighted average of all observation "layers" that overlap the model layer, preserving water content. This process has the advantage that the final observed time series is on each land surface model's vertical coordinate, facilitating comparison.

Where there are multiple instruments at the same station in a network in ISMN, the data are sorted based on the number of days without missing data so that the most complete time series can be accessed easily. Data are gathered by network so that analyses and comparisons can be performed on a network-by-network basis.

An initial concern was whether the differences in the processing of data from the same network taken from each data collection would affect the results, particularly the fact that some gap-filling had been applied to the NASMD time series, but not to ISMN. Furthermore, it is evident from a comparison of Tables 2 and 3 that the number of stations and period of data collected is not the same between the two collections for the same network. Comparison of network statistics, some of which

149 are shown in Section 4, suggest there is no significant difference between the two
150 versions of data for the same networks. Finally, due to varying data availability, for
151 each calculation a fixed period was identified subjectively for each network where
152 most stations have data. Within that period, a station is removed if more than half
153 the period is outside the range of data for that station.

154 *2.4 Models*

155 Four LSMs are confronted with the observational data from ISMN and NASMD: the
156 Catchment model from the National Aeronautics and Space Administration (NASA)
157 Goddard Space Flight Center (GSFC; Koster et al. 2000, Ducharne et al. 2000), the
158 Noah model version 2.7 from the National Oceanic and Atmospheric Administration
159 (NOAA; Ek et al. 2003), the Hydrology-Tiled European Centre for Medium-range
160 Weather Forecasts (ECMWF) Surface Scheme for Exchange over Land (HTESSEL;
161 Balsamo et al. 2009), and the Community Land Model version 4.0 that is sponsored
162 by NSF and DOE (CLM4; Lawrence et al. 2011). Catchment parameterizes an
163 idealized hillslope in each grid box to estimate from bulk water prognostic variables
164 the fractional areas of saturated, unstressed and dry surfaces with respect to
165 evapotranspiration; it then calculates soil moisture profiles as a diagnostic. The
166 other three LSMs calculate soil moisture in each layer as a balance between
167 gravitational drainage and down-gradient conduction in the vertical only. All models
168 treat infiltration of precipitation as a water input at the top, direct evaporation from
169 the top soil layer as an output to the atmosphere, transpiration drawing water out of
170 all soil layers containing roots, and baseflow drainage removing water from the
171 bottom of the soil column. CLM includes deep interaction with a water table
172 parameterization below the soil column. Each model uses its own distributed global
173 map of soil properties on the model grid to determine saturated hydraulic

conductivity, porosity, and other necessary hydraulic parameters. None of the models as used here consider vertical variations in basic soil properties.

Multiple sets of soil moisture data from each of four modeling centers above have been collected and compared. For contributions from a given modeling center, the LSM used is nearly or exactly the same, but the way time series of soil moisture have been produced varies. The contributions from each modeling center include an offline simulation with the LSM driven by gridded global observationally-based meteorological analyses, and a simulation or set of simulations with the LSM coupled to its corresponding global atmospheric model in a free-running (unconstrained or forecast) mode. In the case of Noah and CLM, an ocean general circulation model was also coupled to the atmospheric model but that has little consequence for this study, and for CLM predicted vegetation phenology was enabled. For all but CLM, there is also a reanalysis where the atmosphere and the land surface states, to varying extents, are constrained by data assimilation. Noah is the LSM in two reanalyses investigated here. Table 4 outlines the various configurations and the spatial resolutions of these models. When compared to observed data, the model grid box containing the site of observed station is used.

3. Metrics

The model simulations described in Section 2.4 are confronted with three metrics from the observed soil moisture networks' data. Additionally, the observations themselves are evaluated to assess their likely measurement error based on the methodology of Vinnikov et al. (1996), as described in the next section.

197 The first metric assessed is the variance or standard deviation of daily soil moisture
198 for each month, grouped by season (DJF, MAM, JJA and SON). No attempt is made to
199 remove a climatological annual cycle, as the in situ networks have, by and large, not
200 been in place long enough to calculate stable climatological mean annual cycles for
201 most stations. As we are concerned with linkages between land and atmosphere at
202 sub-seasonal time scales, the mean of each month is removed from all data in that
203 month so that no interannual variability enters the calculation, but some seasonal
204 trends within months may still be present that may affect statistics.

205 Second, the soil moisture memory is assessed for each station and vertical level in
206 the soil by computing lagged autocorrelations of the daily time series. Lagged
207 autocorrelations indicate soil moisture behaves as a first-order Markov process
208 (Schlosser and Milly 2002). As a result, we can estimate time scales of correlation,
209 i.e., memory, as the time it takes the lagged autocorrelation of soil moisture to drop
210 to $1/e$. We have found that linear extrapolation between the values of $\ln(r)$ where r
211 is autocorrelation at lags of 1 and 2 days to the lag where $\ln(r)=-1$ provides an
212 estimate that is not significantly different from using a linear fit through $\ln(r)$ at a
213 larger number of lags (cfr. Robock et al. 1995). This is also calculated on a seasonal
214 basis as there is a pronounced annual cycle of memory time scales in most locations.
215 These are then compared to model grid boxes at the same locations. Third, we
216 perform a similar calculation between time series from pairs of stations, and
217 between their corresponding model grid boxes, to assess spatial correlation and
218 length scales.

219 Finally, we are concerned about the representativeness of point soil moisture
220 measurements for model grid box averages. There are inherent problems with
221 direct comparison between point observations and LSM grid box output (cf. Gruber

et al. 2013, Dirmeyer et al. 2013). The densities of some of the in situ networks are high enough to allow us to assess the sensitivity of an area-average soil wetness at model grid scales to the number of measurements contributing to the average. Thus we attempt to address the issue of scale mismatch and account for it when comparing models and observations. We address the scaling issue in Section 5 before showing model performance on the metrics listed above.

4. Observational Error

Vinnikov and Yeserkepova (1991), following the proposition of Delworth and Manabe (1988), showed that soil moisture time series behave like first-order Markov processes such that the autocorrelation of soil moisture at a location at lag decreases as lag grows:

$$r(\tau) = \exp(-\lambda\tau) \tag{1}$$

where λ is the decay frequency, or $1/\lambda$ is the time scale. Robock et al. (1995) showed that for actual observed data, a linear best fit of ln(r) versus τ for a range of lags does not cross at a value of r=1 (i.e., $\ln(r) = 0$), but rather at some correlation $r < 1$. The displacement of the correlation at $\tau = 0$; a (i.e., $r(\tau = 0) = 1 - a$), is an indicator of measurement error.

Vinnikov et al. (1996) noted that the variance in any time series of observed measurements is composed of the sum of the actual variance of the measured quantity and the noise contributed from random observational error. The ratio of error variance δ^2 to real variance σ^2 is related to the displacement of the extrapolated autocorrelation:

$$\frac{\delta^2}{\sigma^2} = \left(\frac{a}{1+a} \right) \quad (2)$$

This relative error can also be derived from a different perspective using triple colocation (Gruber et al. 2016). Given a sufficient number of measurements, observational error from a station or network of stations can be estimated without specific validation or comparison to independent data. This is a very powerful result that can provide a measure of uncertainty for data that have red noise spectra, such as soil moisture.

We have applied this approach to estimate the error in the networks represented in the ISMN and NASMD data sets. Figure 2 shows the relative random error as the square root of the ratio in Eq. 2 estimated across all stations in each network for the two databases. The data from all networks is interpolated to the four Noah model layers: 0-10cm, 10-40cm, 40-100cm and 100-200cm. Recall that for the same networks ISMN and NASMD do not always contain the same stations or span of years. As a result, the estimated observational errors for the same networks in the two databases do not match exactly – the differences may be taken as representative of the uncertainties in applying this method.

Certain features are apparent nevertheless. As found by Gruber et al. (2013) observational errors are generally largest in the surface layer and decrease with depth. There is also a distinct difference between networks, and in fact between types of instrumentation. The GPS reflection method of the PBO-H2O network appears to result in a large relative random error of measurement of 0.35 for surface soil moisture. Relative random error in the cosmic ray neutron method of the COSMOS network is nearly as large at 0.32. Some of this may not be truly *random* error but rather due to the fact that the effective measurement depth varies

with soil moisture content, so the static station measurement depths used here introduce additional error.

Dielectric probes are inexpensive and thus the most widely used. They have a relative random error of about 0.18 for near surface measurements, dropping to 0.12 below 1m depth. Heat dissipation instruments appear to be the most accurate, with a surface relative random error of 0.15 dropping to around 0.07 at depth. There is a great deal of variation among networks using the same class of instrumentation. For the CHILI network, which places dielectric instruments only at 1m depth, random error appears exceptionally large. The SOILSCAPE network also appears to have unusually large random errors.

A further aspect of this approach to error estimation is that we can define representative profiles of a for classes of instruments, networks or stations, allowing us to estimate a "corrected" soil moisture memory for comparison to models, which by their nature do not suffer from random error in their reported state variables. This can be accomplished by shifting the linear best fit of $\ln(r)$ by a so as to intersect 0 at $\tau = 0$; the corrected estimate of memory from observations is then the lag at which the adjusted $\ln(r) = -1$.

5. Spatial consistency between models and observations

The high-density SoilScape network contains several sets of instruments or "nodes" clustered in groups within areas of <1 km^2 in several locations. At these locations it may be assumed that the stations are so close together that their separations are well within the meteorological spatial scale over which precipitation time series decorrelate. Variations in soil moisture time series from node to node should be due

293 to variations in the hydrologic properties of the sites of each node, and random
294 measurement error.

295 There are also parts of the US where SCAN and SNOTEL have multiple stations
296 separated by distances larger than SoilSCAPE, but within a typical global climate
297 model grid box *O*(100km). Variations on these scales may begin to be determined by
298 differences in the meteorological time series at each station that are not represented
299 explicitly by global models because they are at the sub-grid scale. Thus, data from
300 these sites may allow us to see how this bridging of scales affects the
301 representativeness of climate model soil moisture time series, and shed light on
302 how to compare models to observations across scales.

303 First we consider how averaging together the time series of multiple proximate
304 stations affects the statistics of the time series. The assumption is that a model grid
305 box time series is more representative of the average of multiple stations within
306 that grid box than single stations. If statistics are found to converge as more stations
307 are added to the average, and a general scaling factor is found to apply, we may have
308 a means to translate single-station statistics to model grid box statistics and vice
309 versa. If the statistics do not appear to be sensitive to the number of stations
310 included in the average, then scaling may not be necessary to compare point
311 measurements to model grid box values.

312 We examine data from four of the SoilSCAPE sites that include the most nodes and
313 longest time series. These are located near Canton, Oklahoma (nodes numbered in
314 the 100s), Tonzi Ranch, California (400s), and New Hogan Lake, California (500s and
315 700s). Across all sites the maximum separation of any two nodes is 503m, and the
316 median separation ranges from 166m to 214m. Data span three years from 2011 to
317 2014. We include 19 nodes from the 400s site, 18 from the 500s, 21 from the 100s

318 and 13 from the 700s. Also examined are data from a cluster of 12 SCAN sites in and
319 around northern Alabama that span 2002-2014 and range in separation from 6-
320 118km with a median separation of 52km.

321 Calculations were performed using data vertically interpolated to the Noah model
322 levels – results shown here are confined to the top 10cm layer results, but lower
323 level results are consistent with these. Also we examined data for the entire year,
324 and separately by season. We find the mean and standard deviation in time for time
325 series from each node at a site, then for each combination of two nodes averaged
326 together only for days when each has no missing data, then for combinations of
327 three, four, etc., up to the series where all nodes at a site are averaged together. We
328 also calculate the variance among each statistic calculated with the same number of
329 nodes in the combination. Because of missing data for different dates at various
330 nodes, the total amount of data in the calculations dwindles as we go to larger and
331 larger combinations. Fig 3a shows how data completeness drops from nodes
332 considered one at a time through larger and larger combinations. We also
333 constructed an abbreviated complete time series by taking a subset of the soil
334 moisture data from SCAN stations in and around northern Alabama – ten stations
335 that have complete data from 22 March 2007 through 21 January 2008. This is used
336 to compare the impact of missing data on statistics.

337 Fig 3b shows how the average standard deviation in time for daily surface soil
338 moisture from all seasons changes as more nodes are averaged together. Aside from
339 a small uptick when going from single stations to combinations of two, there is little
340 systematic change and the curves are remarkably flat. The complete data suggest a
341 slight drop in the average standard deviation of ~10% from nodes considered
342 individually to all considered together. Station data during summer only is more apt

343 to show a slight rise in standard deviations with more combined nodes, whereas it is
344 flatter in spring and fall (not shown). Overall, it would seem that the variability of
345 daily soil moisture time series is not very sensitive to spatial scaling and aggregation
346 over *O*(100m) to *O*(<100km), and it appears model gridbox soil moisture can be
347 reasonably validated against single-site data for this metric.

348 This is heartening, since there is certainly sensitivity of the time mean to
349 aggregation of nodes. Fig 3c shows the coefficient of variation (COV) of mean soil
350 moisture across the various numbers of combinations of nodes. There is a general
351 drop in COV as more nodes are included in the averaging. The flattening of the slope
352 of the curve around the middle values of combined nodes followed by an inflection
353 and steepening of the curves again as nearly all nodes are included appears not to be
354 solely an artifact of the drop in data completeness, although it appears to be less
355 pronounced in the complete data subset.

356 Soil moisture memory as defined in Section 3 is also examined. Missing data affect
357 the estimation of memory as lagged autocorrelations can only be estimated when
358 data are not missing on consecutive days (or two days apart for lag-2 estimates). As
359 we average more stations together, there are more days when at least one station is
360 missing data and the sample size decreases more steeply than for mean or variance.

361 Figure 4 shows estimates of top 10cm soil moisture memory as a function of the
362 number of stations combined for four SoilScape locations, SCAN stations over
363 northern Alabama and the complete subset of those same SCAN stations. There are
364 two curves for each set of stations. The solid curves show the median value of soil
365 moisture memory calculated separately across all combinations of stations taken N
366 at a time, N indicated on the abscissa. The dotted line is the memory calculated from
367 the average 1-day and 2-day lagged autocorrelations across all combinations.

There is a clear separation between the three California sites and those east of the Rockies. The California sites have the longer memories, which is logical as California has a prolonged dry season when soil moisture anomalies may persist for months relative to the climatological annual cycle (though we caution that some of the memory may reflect sub-monthly facets of the climatological seasonal cycle, which were not removed by the aforementioned subtraction of monthly means from the data). They also show discrepancies between the two approaches to estimating memory for small numbers of combined stations. The estimates converge in all cases for greater numbers of combined stations. The Oklahoma and Alabama sites have shorter memories, consistent with their year-round likelihood for precipitation and general lack of a dry season. They also show very high agreement between the median memory and the memory calculated from the mean lagged autocorrelations. In all cases the memory calculated from the mean lagged autocorrelations increases as more stations are averaged together – by 14% for the SCAN Alabama sites to nearly 200% for the SoilScape 700 nodes. Behavior of the medians is less consistent, as values go up or down as the number of combined stations increases.

Across this limited number of sites there is not an obvious relationship between characteristics of the station data (e.g., completeness shown in Fig 3a) and the type or degree of change from single stations to inclusion of all stations. It would not be advisable based on these results to propose a method to scale memory calculated from individual stations to grid-box averages. It may be more judicious to compare a number of stations averaged over a certain area to model output averaged over the same area, although as shown later there are suspicious systematic differences between observational networks as well.

6. Variance and Memory

In this section we begin comparing model estimates of soil moisture statistics across the conterminous US with statistics from co-located in situ measurements. For each model configuration, stations are compared to the model grid box that contains them. Table 5 shows the correlations calculated for each of the 19 networks listed in Table 1 between model and observed intra-seasonal (calculated monthly and averaged for seasons) standard deviations of daily soil moisture. Correlations are then averaged across the networks, weighted by the number of stations in each network that went into the calculation. Because of the varying numbers of stations and areal extents of the different networks, it is not feasible to assign statistical significance to the averaged correlations. Networks are separated into local and regional extents because we have noticed a rather systematic separation in the correlations: uniformly the models verify poorly with the local (mostly state-level) networks in terms of the spatial pattern of soil moisture variability, but verify relatively well with the regional and national networks. This suggests that patterns of variability driven by the varying climate regimes across the US are somewhat well reflected by the models, but smaller scale variations over a few hundred kilometers or less are not captured. These smaller scale patterns are likely more determined by variations in soils and landscape that are poorly represented by LSMs at grid-box scales.

Other patterns are evident in Table 5. Correlations between models and observations are generally highest for the shallow layer below the surface layer, which ranges in thickness from ~2cm (Catchment and CLM) to 10cm (Noah). There is no indication that the model output where near-surface meteorology is dictated by observations (offline and reanalysis) is better than free-running models. This is

somewhat surprising, as there are acknowledged problems with global model simulations of precipitation, and precipitation quality is the major control on soil moisture variations (Guo et al. 2006; Wei et al. 2008). The implementation of HTESSEL in ERA-Interim used a single loamy soil texture globally, which may mute discrepancies due to soil property disagreement with observational sites and greater resemblance to forcing (precipitation) patterns thus increasing correlation. Lastly, the various configurations of the Catchment LSM represent this particular metric of soil moisture variability more poorly than any of their counterparts. We should note, however, that soil moisture products from the Catchment LSM have been tested extensively against in situ measurements in other studies (Liu et al. 2011, de Lannoy et al. 2014) and, in terms of capturing the time variability of soil moisture variation at a wide variety of sites, the model performs better (average time correlations of about 0.5) than suggested by the present metric, which focuses instead on the spatial correlation against observations of the temporal standard deviation.

Figure 5 displays the network means and model biases in the daily standard deviation of JJA surface volumetric soil moisture (top 3 layers for CLM) for each model configuration and network in a color-coded tabular form. There are clear systematic biases in the representation of soil moisture variability among the models. Offline versions of both HTESSEL (ERA-I land) and CLM exhibit excessive variance of soil moisture. For the Noah LSM the offline version (GLDAS) also has the highest variance, but still has a low overall bias across all networks. The reanalyses tend to exhibit the lowest variability, although CFSR is essentially undistinguishable from the offline or free-running (CFS) simulations. Among the four model groups,

442 CLM has the strongest positive biases and Catchment (MERRA and GEOS5) has the
443 strongest negative biases.

444 The vagaries of validation with multiple observational networks are also evident
445 when one compares the different rows of Fig 5. Models exhibit the strongest positive
446 biases for the two networks that employ heat-dissipation sensors: ARM and OK-
447 Meso, suggesting these sensors may behave differently than other types of
448 measurements. All models show negative biases for the Missouri network, and most
449 also show negative biases for the West Texas network. Meanwhile, biases are
450 generally positive for DEOS; WTX-Meso and DEOS employ the same model of
451 dielectric soil moisture probe at the same depth for surface soil moisture (5cm), so
452 the systematic differences are more likely due to disparities between the gridded
453 soil parameter data sets commonly used by models and actual local conditions.

454 Soil moisture memory is calculated as described in Section 3. The extrapolation
455 procedure from data during a season can result in e-folding time scales for lagged
456 auto-correlations that vary over two or more orders of magnitude in some cases.
457 Accuracy of long-memory estimates is particularly tenuous, so all averaging is done
458 in terms of frequency rather than time, and the inverse of the result is taken to give
459 a memory time scale in days (i.e., the harmonic mean is used).

460 Table 6 shows how well the spatial patterns of soil moisture memory agree between
461 models and observations. As in Table 5, the network results are grouped by extent:
462 local versus regional/national. As with soil moisture variance, models represent
463 large-scale patterns of memory better than intrastate variations. However,
464 correlations are generally lower for memory than for the standard deviation of soil
465 moisture. The highest skill is exhibited for surface soil moisture memory among the
466 local networks, and for shallow (approx. 10-50cm depth) soil moisture memory at

467 larger scales. Free-running land-atmosphere models perform worst at simulating
468 large-scale patterns of soil moisture memory – this could be due to errors in the
469 temporal spectrum of precipitation in models (cf. Wei et al. 2010; Dirmeyer 2013).
470 Interestingly, free-running models do the best at representing local network
471 variations, but although the correlations are generally statistically significant due to
472 the large number of stations included, they do not suggest practical usefulness.
473 Various configurations with the Noah LSM show the best pattern correlations.

474 Figure 6 presents network-by-network comparisons of JJA surface soil moisture
475 memory biases in the same manner as Fig 5. The mean memory for different
476 networks varies from less than 4 days to more than 17, but there is considerable
477 variation within each network. Model biases can be substantial. Model
478 configurations using the Catchment land surface scheme predominantly exhibit
479 strong positive biases, suggesting that model is overly persistent in soil moisture
480 anomalies, despite the fact it has the thinnest surface layer of the four LSMs. As
481 noted in Sec 2.4, Catchment has a distinctly non-traditional structure; its
482 implementation of an explicit treatment of subgrid soil moisture variability is
483 known to tie together strongly its diagnosed surface and subsurface soil moisture
484 variables (Kumar et al. 2009), which can in fact produce artificially high values of
485 memory in surface and shallow layers. On the other hand, CLM (for which the top 3
486 layers have been combined to represent the surface) largely underestimates soil
487 moisture memory. Noah largely underestimates memory as well, except in the 20th
488 Century Reanalysis where the spatial resolution is considerably coarser than the
489 other implementations, and the large ensemble approach to production (Compo et
490 al. 2011) may have a bearing on hydrologic variables. Another model that shows
491 inconsistent behavior among implementations is HTESSEL – the offline ERA-Interim

Land simulation has consistently negative biases in soil moisture, mirrored also in ECMWF atmospheric coupled integration with data assimilation (Albergel et al. 2012, 2013), while coupled IFS simulations from Athena project (Kinter et al. 2013) and ERA-Interim exhibit a slightly positive bias, highlighting the impact of both modeling and assimilation system changes in determining biases.

Finally, error profiles from network to network are fairly consistent, suggesting again that discrepancies exist between gridded datasets of land surface parameters used by the models and conditions at the station sites. An exception is SNOTEL, in which stations are largely positioned in high mountain locations across the western US that may tend towards thinner and rockier soils than global gridded soil datasets specify or LSMs represent. Catchment has some of its largest positive biases over this network where other models generally have some of their strongest negative biases.

7. Spatial scales

Finally, we examine the spatial decorrelation scales in the observational networks and models. This approach is highly analogous to the temporal scaling we defined as "memory" in Sec 6, and follows similar principles (Vinnikov et al. 1999; Entin et al., 2000). In the case of decorrelation of soil moisture time series over space, we have three factors: decorrelation over meteorological scales of tens to hundreds of kilometers related to the spatial scales of the forcing of soil moisture variability, particularly precipitation; decorrelation over catchment hydrologic scales of meters to hundreds of meters brought about by variations in soil properties and sampling of different regimes along hillslopes; and random measurement errors as characterized in Sec 4.

As a check, we look first at the SOILSCAPE sites, whose nodes are close enough together to be well within the meteorological scales. We find essentially no relationship in any season between the correlation of time series of soil moisture from pairs of stations and their separation distances, which range from about 20-500m. The implication is that evidence for the "catchment hydrologic scale" of Vinnikov et al. (1999) is swamped by the random error in measurements.

For other networks with greater distance between stations, a systematic relationship between station separation and correlation emerges. Figure 7 shows correlation as a function of station separation during summer for several state networks, and larger networks that have numerous stations in a single state ("Delta" refers to all stations in a region centered on northern Mississippi, spanning 93°-85.5°W, 32°-35.1°N; "UT" and "CO" refer to stations within Utah and Colorado). For all networks except the West Texas Mesonet there is a clear decrease in correlation with distance. The two networks using heat dissipation sensors, ARM and the OK-Meso, have the two largest values of extrapolated correlation at distance 0.0, suggesting in a way similar to Fig 2 that these sensors have small random error. They also show higher correlations at larger distances than neighboring AWDN (Nebraska) or other dielectric sensor networks in the Midwest or eastern US. The western networks (SCAN UT, SNOTEL CO and SNOTEL UT) tend to show high correlations at large separations like ARM and OK Meso, likely indicative of the relatively rare precipitation in those areas during summer. The two networks overlapping in Utah (SCAN UT and SNOTEL UT) appear to show very different spatial correlation scales, but when they are adjusted for their different apparent random errors (discussed below), they become quite comparable. SCAN stations are located mainly in agricultural valleys and lowlands while SNOTEL stations are

predominantly in mountains at higher altitudes; it is unclear how this may affect error estimates.

Figure 8 presents a color-coded table of model comparisons to network estimates of the spatial decorrelation distance (km) for surface soil moisture during JJA, defined as the separation between stations where zero-lag temporal correlation of time series of daily soil moisture drops to $1/e$ based on a best fit regression of $\ln(r)$ against distance. For the stations, the lines are shifted so that the intercept at distance 0 is at a correlation of 1, adjusting for the effect of random measurement error on correlations. This has the effect of increasing the spatial decorrelation distance. No such adjustment is necessary for model output, where distances are calculated for each grid box relative to the 8 surrounding grid boxes with distances calculated between the centers of each grid box. Both standard and harmonic means are given across models, and the standard deviation is relative to the standard mean.

Many difficulties in estimating spatial decorrelation scales consistently across networks, and comparing models to them, are evident. The spatial decorrelation distances for each network follow from what was shown in Fig 7. Memories are longer in the west (typically several weeks as opposed to ~1 week in the central and eastern US), and they are long for the two networks with heat dissipation sensors (shaded orange). AWDN (Nebraska) and DEOS (Delaware) show much shorter decorrelation distances than other networks. The contrast between AWDN and both ARM and OK Meso is especially troubling, since they are all close to each other in the Great Plains. Here the different performance of various types of sensors is obvious. Compare to the SCAN-UT and SNOTEL-UT networks over Utah, which use similar

566 instruments and are within 6% of each other once adjusted for random
567 measurement error.

568 Looking at the models, there is a tendency for overestimation of the decorrelation
569 distance. This may be related to the differences in scaling between the point
570 observations and model grid boxes. Even though many pairs of stations are used to
571 estimate the correlations as a function of separation shown in Fig 7, they still
572 represent point-versus-point correlations, and not area-versus-area as is implied in
573 model grid box values. Comparing sets of stations averaged over areas comparable
574 to adjacent model grid boxes might provide a more proportionate estimate, but we
575 immediately face the problem of a lack of station density over all but a few areas. We
576 saw in Section 6 that the estimation of memory, which is a closely related
577 calculation to decorrelation distance (Vinnikov et al. 1996), is not scale resilient like
578 variance.

579 The row of Fig 8 labeled "Corr (no WTX)" is the correlation between model and
580 network estimates across the first 9 networks listed. The curve fit for WTX-Meso is
581 so flat (see Fig 7) that a ridiculously long distance is projected for this network, so it
582 is not included in the correlation. The probability shown is the likelihood the
583 correlation could be arrived at by chance (with 7 degrees of freedom) – a one-tailed
584 test is made as there is no inherent value in a negative correlation. The free-running
585 CFS simulation would appear to perform well, with a correlation of 0.67 and only a
586 5% chance of arriving at this correlation randomly. However, out of 12 models it is
587 not surprising to see one with a probability below 8-10%, and the slope of the fit
588 through the scatter is nearly 1:4, far from 1:1. The harmonic mean of all models
589 produces a lower positive bias than a straight mean, but a negative correlation
590 across the networks.

Among the models there is a great deal of variation. There is not a consistent behavior of offline LSMs relative to coupled models, or reanalyses relative to free-running GCMs. However, the suite of models using the Catchment LSM tends to have shorter decorrelation distances than either the H-TESSEL or CLM sets of models. The Noah set spans a wide range, varying by a factor of two from offline GLDAS to the low-resolution 20CR simulation.

8. Conclusions and Discussion

In this study, we have confronted a number of LSMs in both coupled and uncoupled modes with in situ soil moisture measurements from a number of independent networks over the conterminous United States to (1) determine the feasibility and pitfalls of such a comparison and (2) see what can be learned about model and observational data. We first investigate characteristics of the observational data, with particular attention to how they differ between networks (possibly due to difference in instrumentation), in space, and over time. We then test approaches to compare model output to the observational data.

We examine three statistical metrics: variability as measured by the standard deviation of daily soil moisture, memory represented by the time it takes lagged autocorrelation of daily soil moisture to drop to $1/e$, and spatial scale calculated like memory as the distance over which unlagged correlations between soil moisture measurements or model estimates drop to $1/e$. For measurement networks with many closely-located stations within the area of a typical model grid box, we find that aggregation of many stations (arguably more representative of grid box average values represented by LSMs) has little effect on the standard deviation, but does change estimates of memory in non-systematic ways. Data completeness can affect

aggregation, but not in a clearly predictable way. Although not directly investigated here, there is evidence that spatial scale estimates are also sensitive to the combination of stations. We conclude that modeled soil moisture variability can be safely validated against data from single stations, but other metrics cannot.

Another caveat regarding in situ soil moisture data is that there can be clear differences between the statistical properties of data from different networks that are in the same or adjacent locations. This mirrors what is already established for soil moisture products between models (Koster et al. 2009). In some cases this seems to be caused by the type of sensor used. Buried probes seemed to exhibit less random error than near-field remote sensing techniques, although we do find networks with dielectric sensors that appear to have large random errors. Heat dissipation sensors have generally low random error, but also lower day-to-day variability and connote longer soil moisture memory than dielectric probes, even after random errors are accounted for. Random measurement errors generally decrease with depth for buried sensors. There are also differences in estimated random errors between networks with essentially the same instrumentation, suggesting differences in calibration and maintenance may also be a factor.

Models show systematic biases in near-surface soil moisture metrics (Figs 5, 6, and 8). The model configurations with the Catchment LSM all show too little variability and long memory, but in the case of free-running GEOS5 simulations an unusually short spatial scale in some cases. Bosilovich (2013) showed some precipitation comparisons for MERRA indicating too little interannual variability. CLM tends toward excessive variability and short memory. The characteristics of variability and memory are not always in opposition – ERA-Interim Land has positive biases in both while ERA-Interim has negative in both. These features can be attributed to the

641 modeled soil moisture range lacking spatial variability in ERA-Interim, as
642 documented in the soil hydrology revision (Balsamo et al. 2009). An interesting
643 dependency on spatial resolution can be deduced from the IFS integrations within
644 the Athena project that used a 4-times higher spatial resolution, enhancing the
645 match to in-situ standard deviations of daily surface volumetric soil moisture (Fig.
646 5), memory (Fig. 6), and spatial de-correlation distance (Fig. 8).

647 The various implementations of the Noah LSM range between low and high biases,
648 but often have the lowest biases across all networks. For both variability and
649 memory the models show higher spatial correlations across stations within regional
650 to national networks than to state-level networks (Tables 5 and 6), suggesting they
651 reflect large-scale hydroclimate patterns better than local ones. The specification of
652 soil hydrologic properties in models is much coarser than natural heterogeneity of
653 the soils in most regions, reflecting another aspect of subgrid variability in large-
654 scale models that is poorly represented and confounds validation. Free-running
655 coupled land-atmosphere models perform worst at simulating large-scale patterns,
656 possibly due to the poor simulation of precipitation spectra by GCMs.

657 Spatial scales were found to be particularly difficult to diagnose. For the SoilSCAPE
658 network where stations are separated by only *O*(100m) there appears to be no
659 dependence of inter-station correlation on distance between stations. Theory
660 suggests there is a "catchment hydrologic scale" in this range (Vinnikov et al. 1999)
661 but measurements seem to be dominated by random errors that obscure evidence
662 of it. Meteorological-scale decorrelation over ranges of *O*(10-1000km) are clear in
663 observations and models, but we find the heat dissipation sensors seem to imply
664 longer spatial scales, and different networks of dielectric sensors can give very wide
665 ranges of estimates that appear unrealistic. CLM, Catchment and HTESSEL each

represent subgrid surface variations to differing extents, but this does not translate well to representation of observed subgrid variability (cf. Bosilovich 2002). To put comparisons between models and observations on the same footing for this metric, stations should first be aggregated into average time series over areas comparable to model grid boxes before estimating spatial decorrelation distances.

Overall, statistical vagaries between different soil moisture networks using different types of sensors and measurement techniques suggest great caution is needed when using these data for validation, calibration or data assimilation. The typical assumption that model errors are large while observational errors are small may not apply readily for soil moisture. This is particularly important as the LSM community moves towards a more rigorous benchmarking approach (e.g. Best et al. 2015) for fluxes and other variables such as soil moisture. The results here also suggest statistical considerations that should be applied when extending model evaluation or benchmarking to 2-D rather than at a single point. One must be very careful about scaling issues – everything possible should be done to put data from different sources on the same footing before comparison. Only temporal variability seemed to be insensitive to the differences in scale between point measurements and model grid box values.

Note that this study is largely exploratory and we opt to go wide rather than deep; instead of giving a highly detailed examination of a particular soil moisture metric, observational network or model, we traverse data and metrics from multiple observational networks and models. Our aim is to explore the problems and pitfalls, as well as bring to light the areas of promise for validation of models with observed soil moisture data. The indicated biases potentially indicate deficiencies in the land models. However, they may also reflect the model-specific character of a given land

models' soil moisture representation. Given the necessity of computing fluxes averaged over large and complex domains with limited spatial resolution in soil moisture description, soil moisture in land models is arguably better interpreted as a model-specific index of wetness than a variable that can be directly compared to observations (Dirmeyer 2004, Koster et al. 2009). For this reason, modeled soil moisture is known to have model-specific magnitudes (and yet still function well in climate models); by the same token, standard deviations of soil moisture will also necessarily be model-specific. This point underscores a major difficulty faced when confronting land models with such observations.

The work presents several means to approach the assessment of model soil moisture behavior with in situ observations with particular focus on spatiotemporal inconsistencies. The problem is analogous to that faced in operational data assimilation, where observations from a wide range of sources with different spatiotemporal coverage and error characteristics must be harmonized to generate useful analyses. Key to such approaches is a large and robust set of calibration and validation data – none of the networks examined here are due to be discontinued and more networks are coming online and being synthesized into NASMD and ISMN every year, so the situation should only improve. Furthermore, satellites can provide spatially continuous measurements at scales comparable to model grid boxes. Current missions are beginning to provide such information, but better temporal coverage at higher spatial resolution, maintained uninterrupted over decades will provide, in combination with in situ measurements and synthesis through data assimilation, the best overall monitoring and initialization for forecasts. Despite this, models can still be improved using the growing

observational record to identify and improve processes and parameterizations that contribute to errors in the surface water cycle.

Acknowledgements: This work has been primarily supported by National Aeronautics and Space Administration grant NNX13AQ21G. Funding for WD has come from SMOS Soil Moisture Network Study–Operational Phase (ESA ESTEC Contract No.4000102722/10). Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office – we thank G. Compo for making 20CR data available to us.

728 **References:**

729 Albergel, C., P. de Rosnay, G. Balsamo, L. Isaksen, and J. Muñoz-Sabater, 2012: Soil
730 Moisture Analyses at ECMWF: Evaluation Using Global Ground-Based In Situ
731 Observations. *J. Hydrometeor.*, **13**, 1442–1460. doi:
732 http://dx.doi.org/10.1175/JHM-D-11-0107.1.

733 Albergel, C., W. Dorigo, R. H. Reichle, G. Balsamo, P. de Rosnay, J. Muñoz-Sabater, L.
734 Isaksen, R. de Jeu, and W. Wagner, 2013: Skill and Global Trend Analysis of Soil
735 Moisture from Reanalyses and Microwave Remote Sensing. *J. Hydrometeor.*, **14**,
736 1259–1277. doi: http://dx.doi.org/10.1175/JHM-D-12-0161.1.

737 Andresen J., L. Olse, T. Aichele, B. Bishop, J. Brown, J. Landis, S. Marquie and A.
738 Pollyea, 2011: Enviro-weather: A weather-based pest and crop management
739 information system for Michigan. *7th Int. Integrated Pest Management*
740 *Symposium*, Memphis, TN, 16.1 [Available online at
741 http://www.ipmcenters.org/ipmsymposium12/16-1_Andresen.pdf.].

742 Baldocchi, D., and co-authors, 2001: FLUXNET: A new tool to study the temporal and
743 spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux
744 densities. *Bull. Amer. Meteor. Soc*., **82**, 2415-2434.

745 Balsamo, G., P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A. K. Betts, and K.
746 Scipal, 2009: A revised hydrology for the ECMWF model: Verification from field
747 site to terrestrial water storage and impact in the Integrated Forecast System. *J.*
748 *Hydrometeor*., **10**, 623–643.

749 Balsamo, G., and co-authors, 2015: ERA-Interim/Land: a global land surface
750 reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389-407, doi: 10.5194/hess-19-
751 389-2015.

752 Bell J.E., and co-authors, 2013: U.S. Climate Reference Network soil moisture and
753 temperature observations. *J. Hydrometeor*., **14**, 977-988.

754 Best, M. J., and co-authors, 2015: The plumbing of land surface models:
755 benchmarking model performance. *J. Hydrometeor*., **16**, 1425-1442, doi:
756 10.1175/JHM-D-14-0158.1.

757 Bond D., 2005: *Soil Water and Temperature System (SWATS) Handbook*, Atmospheric
758 Radiation Measurement Climate Research Facility Technical Report TR-063. U.S.
759 Department of Energy, Washington D.C.

760 Bosilovich, M. G., 2002: On the use and validation of mosaic heterogeneity in
761 atmospheric numerical models. *Geophys. Res. Lett*. **29**, 15-1–15-4, doi:
762 10.1029/2001GL013925.

763 Bosilovich, M. G., 2013: Regional climate and variability of NASA MERRA and recent
764 reanalyses: US summertime precipitation and temperature. *J. Appl. Meteor. Clim*.,
765 **52**, 1939-1951, doi: 10.1175/JAMC-D-12-0291.1.

766 Compo, G. P., and co-authors, 2011: The Twentieth Century Reanalysis Project.
767 *Quart. J. Roy. Meteor. Soc*., **137**, 1-28. doi: 10.1002/qj.776.

768 Crow, W. T., F. Lei, M. Anderson, and C. Hain, 2015: Utilizing remote sensing and
769 triple collocation for estimating one-way surface soil moisture and latent heat
770 flux coupling strength. *Geophys. Res. Lett.,* (submitted).

771 de Lannoy, G. J. M., R. D. Koster, R. H. Reichle, S. P. P. Mahanama, and Q. Liu, 2014: An
772 updated treatment of soil texture and associated hydraulic properties in a global
773 land modeling system. *J. Adv. Mod. Earth Syst*., **6**, doi:10.1002/2014MS000330.

774 Dee, D. P., and co-authors, 2011: The ERA-Interim reanalysis: configuration and
775 performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc*., **137**, 553-
776 597, doi:10.1002/qj.828.

777 Dirmeyer, P. A., 2004: Soil moisture - muddy prospects for a clear definition. *GEWEX*
778 *News*, **14**(3), 11-12.

779 Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki and N. Hanasaki, 2006: The Second
780 Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for

781 our perception of the land surface. *Bull. Amer. Meteor. Soc*., **87**, 1381-1397, doi:
782 10.1175/BAMS-87-10-1381.

783 Dirmeyer, P. A., 2013: Characteristics of the water cycle and land-atmosphere
784 interactions from a comprehensive reforecast and reanalysis data set: CFSv2.
785 *Climate Dyn*., **41**, 1083-1097, doi: 10.1007/s00382-013-1866-x.

786 Dirmeyer, P. A., S. Kumar, M. J. Fennessy, E. L. Altshuler, T. DelSole, Z. Guo, B. Cash
787 and D. Straus, 2013: Evolution of land-driven predictability in a changing climate.
788 *J. Climate*, **26**, 8495-8512, doi: 10.1175/JCLI-D-13-00029.1.

789 Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A.,
790 Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T. (2011).
791 The International Soil Moisture Network: a data hosting facility for global in situ
792 soil moisture measurements, *Hydrol. Earth Syst. Sci.,* **15**, 1675-1698, doi:
793 10.5194/hess-15-1675-2011.

794 Dorigo, W.A., and co-authors, 2013: Global automated quality control of in situ soil
795 moisture data from the International Soil Moisture Network. *Vadose Zone J*.,
796 **12**(3), doi: 10.2136/vzj2012.0097.

797 Dorigo, W. A., and co-authors, 2015: Evaluation of the ESA CCI soil moisture product
798 using ground-based observations. *Remote Sens. Env*., **162**, 380-395, doi:
799 10.1016/j.rse.2014.07.023.

800 Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A
801 catchment-based approach to modeling land surface processes in a general
802 circulation model, 2, Parameter extimation and model demonstration. *J. Geophys.*
803 *Res*., **105**, 24823-24838.

804 Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D.
805 Tarplay, 2003: Imple- mentation of Noah land surface model advances in the
806 National Centers for Environmental Prediction operational mesoscale Eta model.
807 *J. Geophys. Res*., **108**, 8851, doi: 10.1029/2002JD003296.

808 Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhai, 2000:
809 Temporal and spatial scales of observed soil moisture variations in the
810 extratropics. *J. Geophys. Res.*, **105**, 11865-11877.

811 Ford, T. W., and S. M. Quiring, 2014a: Comparison and application of multiple
812 methods for temporal interpolation of daily soil moisture. *International J.*
813 *Climatol.*, **34**, 2604-2621, doi: 10.1002/joc.3862.

814 Gruber, A., W. A. Dorigo, S. Zwieback, A. Xaver, and W. Wagner, 2013: Characterizing
815 coarse-scale representativeness of in-situ soil moisture measurements from the
816 International Soil Moisture Network. *Vadose Zone J.*, **12**, doi:
817 10.2136/vzj2012.0170.

818 Gruber, A., C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent
819 advances in (soil moisture) triple collocation analysis. *Int. J. Applied Earth Obs.*
820 *Geoinfo.*, **45**, 200-211, doi: 10.1016/j.jag.2015.09.002.

821 Guinan P. E., and Travlos J. S., 2008: Missouri's transition to a near real-time
822 mesonet. *17th Conf. Applied Climatology,* Whistler, BC, Canada. Amer. Meteor. Soc.,
823 1.2 [Available online at
824 https://ams.confex.com/ams/13MontMet17AP/techprogram/paper_140960.ht
825 m.]

826 Guo, Z., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao, 2006: Evaluation of GSWP-2 soil
827 moisture simulations, Part II: Sensitivity to external meteorological forcing. *J.*
828 *Geophys. Res.*, **111**, D22S03, doi: 10.1029/2006JD007845.

829 Hubbard K.G., Rosenberg N.J., and Nielsen D.C., 1983: Automated weather data
830 network for agriculture. *J. Wat. Resour. Plan. Manag.*, **109**, 213-222.

831 Illston B. G., J. B. Basara, C. A. Fiebrich, K. C. Crawford, E. Hunt, D. K. Fisher, R. Elliott,
832 and K. Humes, 2008: Mesoscale monitoring of soil moisture across a statewide
833 network. *J. Atmos. Oceanic Technol.*, **25**, 167-182.

834 Jackson, T. J. , M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. S. Seyfried, D. C.
835 Goodrich, and M. S. Moran, 2010: Validation of advanced microwave scanning
836 radiometer soil moisture products. *IEEE Trans. Geosci. Rem. Sens.*, **48**, 4256-4272.

837 Kimball, S. K., M. S. Mulekar, S. Cummings, and J. Stamates, 2010: The University of
838 South Alabama Mesonet and Coastal Observing System: A technical and statistical
839 overview. *J. Atmos. Oceanic Technol*., **27**, 1417-1439.

840 Kinter III, J. L., and co-authors, 2013: Revolutionizing climate modeling – Project
841 Athena: A multi-institutional, international collaboration. *Bull. Amer. Meteor. Soc*.,
842 **94**, 231–245, doi: 10.1175/BAMS-D-11-00043.1.

843 Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A
844 catchment-based approach to modeling land surface processes in a general
845 circulation model, 1, Model structure. *J. Geophys. Res*., **105**, 24809-24822.

846 Koster, R. D., Z. Guo, P. A. Dirmeyer, R. Yang, K. Mitchell, and M. J. Puma, 2009: On the
847 nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, doi:
848 10.1175/2009JCLI2832.1.

849 Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009:
850 Role of subsurface physics in the assimilation of surface soil moisture
851 observations. *J. Hydrometeor.*, **10**, 1534-1547.

852 Larson, K. M., E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny,
853 2008: Use of GPS receivers as a soil moisture network for water cycle studies.
854 *Geophys. Res. Lett*., **35**, L24405, doi: 10.1029/2008GL036013.

855 Lawrence, D. M., and co-authors, 2011: Parameterization improvements and
856 functional and structural advances in version 4 of the Community Land Model. *J.*
857 *Adv. Model. Earth Syst*., **3**, doi: 10.1029/2011MS000045.

858 Legates, D. R., D. J. Leathers, T. L. DeLiberty, G. E. Quelch, K. Brinson, J. Butke, R.
859 Mahmood, and S. A. Foster, 2005: DEOS: The Delaware Environmental Observing
860 System. *21st Int. Conf. Interactive Information Processing Systems (IIPS)*, San

861 Diego, CA, Amer. Meteor. Soc., 18.10 [Available online at
862 https://ams.confex.com/ams/Annual2005/techprogram/paper_87687.htm.]

863 Legates, D. R., D. J. Leathers, T. L. DeLiberty, G. E. Quelch, and K. Brinson, 2007:
864 Delaware Environmental Observing System: An Update. *23rd Int. Conf. Interactive*
865 *Information Processing Systems (IIPS)*, San Antonio, TX. Amer. Meteor. Soc., P1.2
866 [Available online at
867 https://ams.confex.com/ams/87ANNUAL/techprogram/paper_120626.htm.]

868 Liu, Q., and co-authors, 2011: The contributions of precipitation and soil moisture
869 observations to the skill of soil moisture estimates in a land data assimilation
870 system. *J. Hydromet.*, **12**, 750-765.

871 Moghaddam, M., D. Entekhabi, Y. Goykhman, K. Li, M. Liu, A. Mahajan, A. Nayyar, D.
872 Shuman, and D. Teneketzis, 2010: A wireless soil moisture smart sensor web
873 using physics-based optimal control: concept and initial demonstration. *IEEE-*
874 *JSTARS*, **3**, 522-535, doi: 10.1109/JSTARS.2010.2052918.

875 Pan W., Boyles R. P., White J. G., and Heitman J. L., 2012: Characterizing soil physical
876 properties for soil moisture monitoring with the North Carolina Environment and
877 Climate Observing Network. *J. Atmos. Oceanic Technol*., **29**, 933-943.

878 Qian, T., A. Dai, and K. E. Trenberth, 2007: Hydroclimatic trends in the Mississippi
879 River Basin from 1948 to 2004. *J. Climate*, **20**, 4599–4614.

880 Quiring, S. M., T. W. Ford, J. K. Wang, A. Khong, E. Harris, T. Lindgren, D. W. Goldberg,
881 and Z. Li, 2015: North American Soil Moisture Database: Development and
882 applications. *Bull. Amer. Meteor. Soc*., (in review).

883 Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama,
884 and A. Touré, 2011: Assessment and enhancement of MERRA land surface
885 hydrology estimates. *J. Climate*, **24**, 6322–6338, doi: 10.1175/JCLI-D-10-05033.1.

886 Rienecker, M. M., and co-authors, 2011: MERRA: NASA's Modern-Era Retrospective
887 Analysis for Research and Applications. *J. Climate*, **24**, 3624-3648,
888 doi:10.1175/JCLI-D-11-00015.1.

889 Robock, A., K. Ya. Vinnikov, C. A. Schlosser, N. A. Speranskaya and Y. Xue, 1995: Use
890 of midlatitude soil moisture and meteorological observations to validate soil
891 moisture simulations with biosphere and bucket models.. *J. Climate*, **8**, 15-35.

892 Rodell, M., and co-authors, 2004. The Global Land Data Assimilation System, *Bull.*
893 *Amer. Meteor. Soc*., **85**, 381-394.

894 Saha, S. and co-authors, 2010: The NCEP climate forecast system reanalysis. *Bull.*
895 *Amer. Meteor. Soc*., **91**, 1015–1057, doi: 10.1175/2010BAMS3001.1.

896 Schaefer G.L., and Paetzold R.F., 2001: SNOTEL (SNOwpack TELemetry) and SCAN
897 (Soil Climate Analysis Network). *Proc. Intl. Workshop on Automated Wea. Stations*
898 *for Appl. in Agr. and Water Resour. Mgmt.*, Lincoln, Nebraska, High Plains Climate
899 Center and WMO, AGM-3, WMO/TD No. 1074.

900 Schaefer G. L., M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources
901 Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Oceanic*
902 *Technol*., **24**, 2073-2077.

903 Schlosser, C. A., and P. C. D. Milly, 2002: A model-based investigation of soil moisture
904 predictability and associated climate predictability. *J. Hydrometeor*., **3**, 483-501.

905 Schroeder J.L., Burgett W.S., Haynie K.B., Sonmez I., Skwira G.D., Doggett A.L., and
906 Lipe J.W., 2005: The West Texas Mesonet: A technical overview. *J. Atmos. Oceanic*
907 *Technol*., **22**, 211-222.

908 Stillman, S., J. Ninneman, X. Zeng, T. Franz, R. L. Scott, W. J. Shuttleworth, and K.
909 Cummins, 2014: Summer soil moisture spatiotemporal variability in
910 southeastern Arizona. *J. Hydrometeor*., **15**, 1473-1485; doi: 10.1175/JHM-D-13-
911 0173.1.

912 Vinnikov, K. Ya., and I. B. Yeserkepova, 1991: Soil moisture, empirical data and
913 model results. *J. Climate*, **4**, 66-79.

914 Vinnikov, K. Ya., A. Robock, N. A. Speranskaya, and C. A. Schlosser, 1996: Scales of
915 temporal and spatial variability of midlatitude soil moisture at different levels. *J.*
916 *Geophys. Res*., **101**, 7163-7174.

917 Vinnikov, K. Ya., A. Robock, S. Qiu, and J. K. Entin, 1999: Optimal design of surface
918 networks for observation of soil moisture. *J. Geophys. Res*., **104**, 19743-19749.

919 Wei, J., P. A. Dirmeyer, and Z. Guo, 2008: Sensitivities of soil wetness simulation to
920 uncertainties in precipitation and radiation, *Geophys. Res. Lett*. **35**, L15703, doi:
921 10.1029/2008GL034494.

922 Wei, J., P. A. Dirmeyer, and Z. Guo, 2010: How much do different land models matter
923 for climate simulation? Part II: A temporal decomposition of land-atmosphere
924 coupling strength. *J. Climate*, **23**, 3135-3145.

925 Xia, Y., M. Ek, Y. Wu, T. Ford, and S. Quiring, 2015: Comparison of NLDAS-2
926 Simulated and NASMD Observed Daily Soil Moisture. Part I: Comparison and
927 Analysis. *J. Hydrometeor*., (early release), doi: 10.1175/JHM-D-14-0096.1.

928 Zamora R. J., F. M. Ralph, E. Clark, and T. Schneider, 2011: The NOAA
929 Hydrometeorology Testbed soil moisture observation networks: Design,
930 instrumentation, and preliminary results. *J. Atmos. Oceanic Technol*., **28**, 1129-
931 1140.

932 Zreda, M., W. J. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz, and R. Rosolem,
933 2012: COSMOS: The COsmic-ray Soil Moisture Observing System, *Hydrol. Earth*
934 *Syst. Sci.,* **16**, 4079-4099, doi: 10.5194/hess-16-4079-2012.

935 Table 1. Observational networks used in this study.

Collection	Network	Full Name	Location	Probe type(s)	Reference(s)
NASMD	AMERIFLUX	AmeriFlux	U.S.	Various Dielectric	Baldocchi et al. 2011
Both	ARM	Dept. of Energy Atmospheric Radiation Measurement	Oklahoma, Kansas	Campbell 229L, SMP1	Bond 2005
Both	AWDN	Automated Weather Data Network	Nebraska	Vitel, ThetaProbe ML2X	Hubbard et al. 1983
NASMD	CHILI	Center for Hurricane Intensity and Landfall Investigation	Alabama	Hydraprobe	Kimball et al. 2010
Both	COSMOS	Cosmic Ray Soil Moisture Observing Station	U.S	Cosmic-ray Probe	Zreda et al. 2012
NASMD	DEOS	Delaware Environmental Observing System	Delaware	Campbell 616L	Legates et al. 2005, 2007
NASMD	ECONET	North Carolina Environment and Climate Observing Network	North Carolina	ThetaProbe	Pan et al. 2012
NASMD	MAWN	Michigan Automated Weather Network	Michigan	Campbell 616	Andresen et al. 2011
NASMD	MAW-MO	Missouri Agricultural Electronic Bulletin Board	Missouri	Campbell 616	Guinan and Travlos 2008
NASMD	NOAAHMT	NOAA Hydrometeorology Testbed Observing Network	U.S.	Campbell 616 & Stevens Hydra Probe	Zamora et al. 2011
NASMD	OK-MESO	Oklahoma Mesonet	Oklahoma	Campbell 229L	Illston et al. 2008
ISMN	PBO-H2O	Xenon PBO H_2O	Western U.S.	GPS	Larson et al. 2008
Both	SCAN	USDA Soil Climate Analysis Network	U.S.	Hydraprobe (Analog and digital)	Schaefer et al. 2007
NASMD	SDAWN	South Dakota Automated Weather Data Network	South Dakota	Stevens Hydra Probe	–
Both	SNOTEL	Snowpack Telemetry Network	Western U.S.	Hydraprobe (Analog and digital)	Schaefer and Paetzold 2001
ISMN	SOILSCAPE	SoilSCAPE	U.S.	EC5	Moghaddam et al. 2010
Both	USCRN	NOAA U.S. Climate Reference Network	U.S.	Stevens Hydraprobe II Sdi-12	Bell et al. 2013
ISMN	USDA-ARS	USDA Agricultural Research Service	U.S.	Hydraprobe Analog	Jackson et al. 2010
NASMD	WTX-MESO	West Texas Mesonet	Texas	Campbell 616L	Schroeder et al. 2005

Table 2. Details of ISMN networks used in this study. Local extent means the network spans parts of one (or for ARM, two) states; regional networks span many states, and often all of the conterminous US. Interval indicates typical interval – for some networks labeled "Hourly" some subset of instruments may report 3- or 6-hourly. Years and days indicate the smallest span that incorporates all data. Instruments indicates the sum of the maximums of the number of each unique type or uniquely labeled type of sensor deployed at any site. SM levels indicates the total number of unique depths of sensor placement for the network; * indicates sensors that sit above ground and sense soil moisture remotely – the depth of the measurement is not static and varies with location, soil moisture and other conditions.

Network	Extent	Interval	Stations	Years	Days	SM Levels	Instruments
ARM	Local	Hourly	29	22	7740	10	15
AWDN	Local	Daily	50	13	4749	4	2
COSMOS	Regional	Hourly	101	7	2308	1*	1
PBO-H2O	Regional	Daily	108	8	2865	1*	3
SCAN	Regional	Hourly	211	19	6877	24	52
SNOTEL	Regional	Hourly	415	19	6425	16	45
SOILSCAPE	Regional	Hourly	135	4	7256	29	6
USCRN	Regional	Hourly	114	15	5029	5	5
USDA-ARS	Regional	Hourly	4	8	2618	1	4

Table 3. As in Table 2 for NASMD.

Network	Extent	Interval	Stations	Years	Days	SM Levels	Instruments
AMERIFLUX	Regional	Daily	55	13	5844	37	1
ARM	Local	Daily	17	16	5845	10	1
AWDN	Local	Daily	41	5	1827	4	1
CHILI	Local	Daily	25	3	1097	1	1
COSMOS	Regional	Daily	54	5	1828	1*	1
DEOS	Local	Daily	26	12	3289	1	1
ECONET	Local	Daily	31	15	5479	1	1
MAWN	Local	Daily	80	17	6209	2	1
MAW-MO	Local	Daily	8	12	4018	1	1
NOAAHMT	Regional	Daily	25	16	4384	3	1
OK-MESO	Local	Daily	104	13	4749	4	1
SCAN	Regional	Daily	123	17	6210	13	1
SNOTEL	Regional	Daily	351	13	4750	16	1
SDAWN	Local	Daily	11	7	2558	5	1
USCRN	Regional	Daily	113	4	1462	27	1
WTX-MESO	Local	Daily	53	11	4018	4	1

952 Table 4. Specifications for the four land and atmosphere models, including source of
953 data and spatial resolution.

LSM	Offline	Free-Running	Reanalysis
Catchment *3 layers* *spanning* *2.01m*	MERRA-Land (MERRA+GPCP forcing) 0.67°x0.5° Reichle et al. (2011)	GEOS5 simulation on MERRA2 mode 0.67°x0.5°	MERRA 0.67°x0.5° Rienecker et al (2011)
Noah2.7 *4 layers* *spanning* *2.00m*	GLDAS 1°x1° Rodell et al. (2004)	CFS seasonal forecasts 0.94°x0.95°	CFSR 0.31°x0.37° Saha et al. (2010); 20CRv2 1.88°x1.91° Campo et al. (2011)
HTESSEL *4 layers* *spanning* *2.89m*	ERA-Interim Land 0.75°x0.75° Balsamo et al. (2015)	IFS in Athena Project 0.14°x0.14° Kinter et al. (2013)	ERA-Interim 0.75°x0.75° Dee et al. (2011)
CLM4.0 *12 layers* *spanning* *3.43m*	Qian et al. (2007) forcing 1.25°x0.9° Lawrence et al. (2011)	CCSM4 seasonal forecasts 1.25°x0.9° Dirmeyer et al. (2013)	--none--

954
955

Table 5. Spatial pattern correlations of subseasonal JJA standard deviations of daily soil moisture between stations and model grid boxes for indicated models. Estimates are grouped by extent of network (local versus regional) and averaged across networks weighted by number of stations used in each network. Surface refers to the top LSM soil level, except CLM where it is the total for the top 3 layers; shallow is the second layer except for CLM where it is layers 4-6; deep is the third layer for Noah and HTESSEL, layers 7-8 for CLM.

LSM	Local				Regional			
Noah	GLDAS	CFSR	CFS	20CR	GLDAS	CFSR	CFS	20CR
Surface	-0.13	0.06	-0.01	0.06	0.38	0.40	0.30	0.25
Shallow	0.08	0.09	-0.06	0.06	0.39	0.34	0.33	0.30
Deep	0.09	0.00	-0.09	0.00	0.25	0.21	0.22	0.18
Catchment	MERRA land	MERRA	GEOS5		MERRA land	MERRA	GEOS5	
Surface	-0.13	-0.11	-0.09		0.08	-0.04	0.21	
Shallow	0.07		0.10		0.29		0.21	
HTESSEL	ERA-In. land	ERA Interim	IFS		ERA-In. land	ERA Interim	IFS	
Surface	0.07	0.03	0.12		0.40	0.50	0.47	
Shallow	0.09	0.03			0.34	0.43		
Deep	-0.02	0.01			0.28	0.32		
CLM	CLM		CCSM		CLM		CCSM	
Surface	0.14		0.07		0.41		0.39	
Shallow	0.11		0.05		0.33		0.34	
Deep	0.07		-0.03		0.27		0.25	
Multi-model	Offline	Reanal.	Free		Offline	Reanal.	Free	
Surface	-0.01	0.01	0.09		0.31	0.28	0.35	
Shallow	0.09	0.06	0.12		0.34	0.36	0.29	
Deep	0.03	0.00	0.04		0.27	0.24	0.22	

965 Table 6. As in Table 5 for soil moisture memory.

LSM	Local				Regional			
Noah	GLDAS	CFSR	CFS	20CR	GLDAS	CFSR	CFS	20CR
Surface	0.11	0.11	0.16	0.17	0.30	0.20	0.18	0.22
Shallow	-0.07	0.05	0.06	0.15	0.41	0.30	0.21	0.21
Deep	-0.03	-0.07	-0.04	-0.01	0.15	0.17	0.12	0.10
Catchment	MERRA land	MERRA	GEOS5		MERRA land	MERRA	GEOS5	
Surface	0.23	0.18	0.20		0.17	0.20	0.16	
Shallow	0.05		0.04		0.04		0.09	
HTESSEL	ERA-In. land	ERA Interim	IFS		ERA-In. land	ERA Interim	IFS	
Surface	-0.09	0.19	0.16		0.14	0.08	0.06	
Shallow	0.09	0.06			0.15	0.25		
Deep	0.02	-0.07			0.10	0.29		
CLM	CLM		CCSM		CLM		CCSM	
Surface	0.08		-0.02		0.15		0.08	
Shallow	-0.03		-0.09		0.25		0.08	
Deep	-0.02		-0.01		0.15		0.02	
Multi-model	Offline	Reanal.	Free		Offline	Reanal.	Free	
Surface	0.08	0.16	0.16		0.19	0.17	0.12	
Shallow	0.01	0.09	0.11		0.21	0.25	0.13	
Deep	-0.01	-0.05	0.05		0.09	0.19	0.07	

966
967

968 **Figure Captions:**

969 Figure 1. Station locations for the networks of (a) ISMN; and (b) NASMD.

970 Figure 2: Estimated relative random observational error as a fraction of estimated
971 real soil moisture variability as a function of depth (interpolated to Noah LSM
972 layers) for each observational network in ISMN (top) and NASMD (bottom). "All" is a
973 weighted average of all networks by the number of stations in each network.

974 Figure 3: Aggregation statistics for observational networks (SS are sets of nodes
975 from SOILSCAPE, 100 is a set of stations in Oklahoma, the others are in California;
976 SCAN AL is a set of SCAN stations across northern Alabama) with closely located
977 stations – values for each number of stations averaged together (abscissa) is the
978 mean for all combinations with the indicated count; (a) number of days of data
979 available when any station with missing data results in missing data for the
980 combined average; (b) the estimated standard deviation in time of daily soil
981 moisture (the mean for each month is first subtracted to remove interannual
982 variability); (c) the coefficient of variation across combinations of the combination
983 means. Black line is from a subset of SCAN AL with no missing data from any station
984 for 306 consecutive days.

985 Figure 4: As in Fig 3 for soil moisture memory (see text for definition). Solid lines
986 are the median value of memory among all combinations, and dotted lines show the
987 harmonic mean across all combinations.

988 Figure 5: Average standard deviation of daily surface volumetric soil moisture
989 (dimensionless) during JJA for stations in each network (first column after network
990 names), and the average bias of standard deviation for each model across the station
991 locations of each network. Averages and standard deviations of each network and

model are shown at the right and bottom of the grid respectively. Color shading is to aid recognition of negative (blue) and positive (red) biases. Two networks that employ heat-dissipation sensors are shaded orange; two networks that use near-field remote sensing techniques are shaded green; all others use some form of dielectric probe.

Figure 6: As in Fig 5 for surface soil moisture memory (units of days).

Figure 7: Spatial autocorrelation of daily surface layer soil moisture during JJA as a function of separation between stations for several networks, binned in intervals of 10km. Lines indicate linear best fits through the binned log of correlation (negative values omitted).

Figure 8: Average spatial decorrelation distance (km) where zero-lag temporal correlation of time series of daily soil moisture drops to 1/e for surface volumetric soil moisture during JJA for stations in each network (first column after network names), and models. Colors indicate percentage deviation of each model (averaged over domain of network) relative to in situ network estimates with estimated random measurement error removed. SD is standard deviation among models.

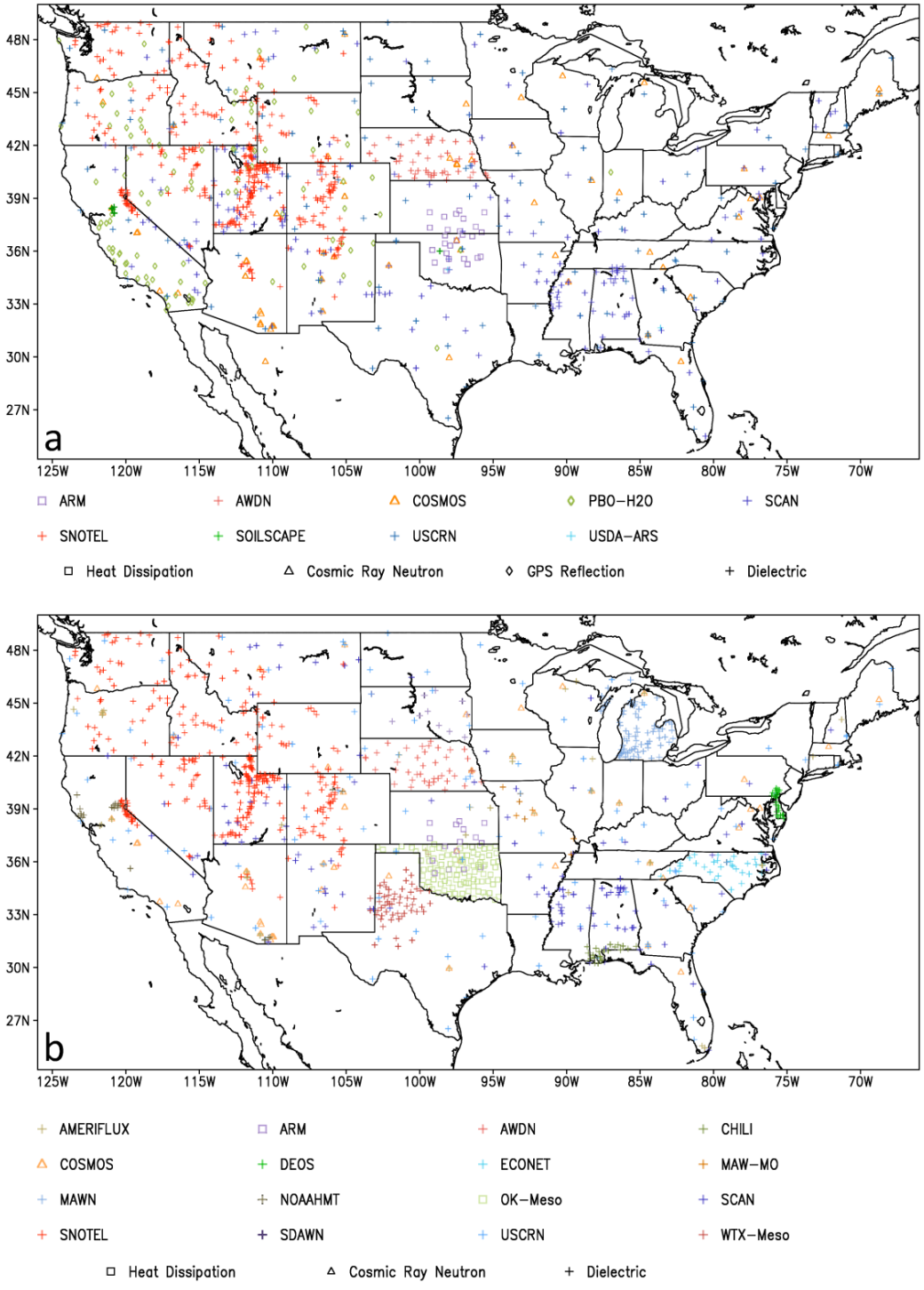

Figure 1. Station locations for the networks of (a) ISMN; and (b) NASMD.

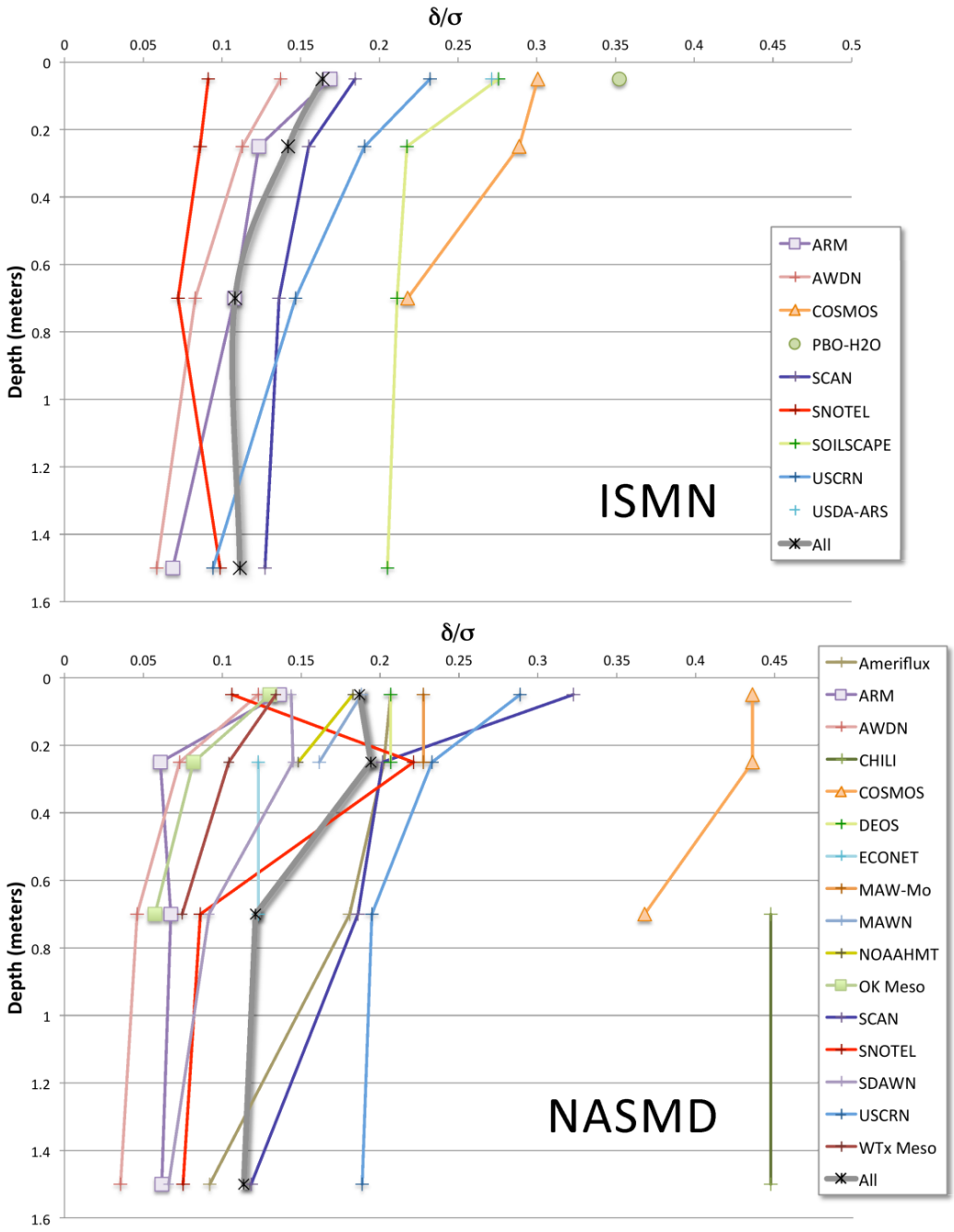

Figure 2: Estimated relative random observational error as a fraction of estimated real soil moisture variability as a function of depth (interpolated to Noah LSM layers) for each observational network in ISMN (top) and NASMD (bottom). "All" is a weighted average of all networks by the number of stations in each network.

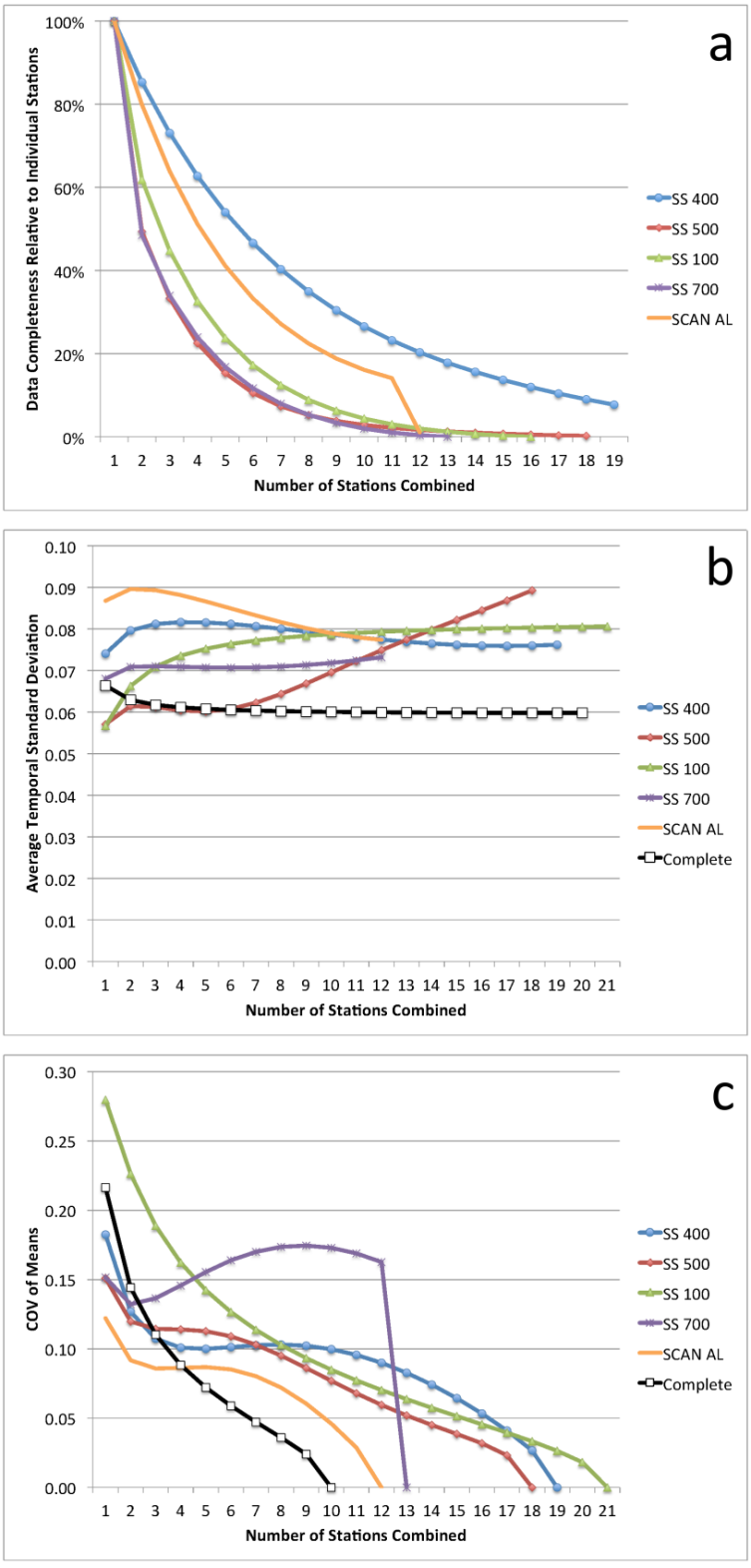

Fig 3: Aggregation statistics for observational networks (SS are sets of nodes from SOILSCAPE, 100 is a set of stations in Oklahoma, the others are in California; SCAN AL is a set of SCAN stations across northern Alabama) with closely located stations – values for each number of stations averaged together (abscissa) is the mean for all combinations with the indicated count; (a) number of days of data available when any station with missing data results in missing data for the combined average; (b) the estimated standard deviation in time of daily soil moisture (the mean for each month is first subtracted to remove interannual variability); (c) the coefficient of variation across combinations of the combination means. Black line is from a subset of SCAN AL with no missing data from any station for 306 consecutive days.

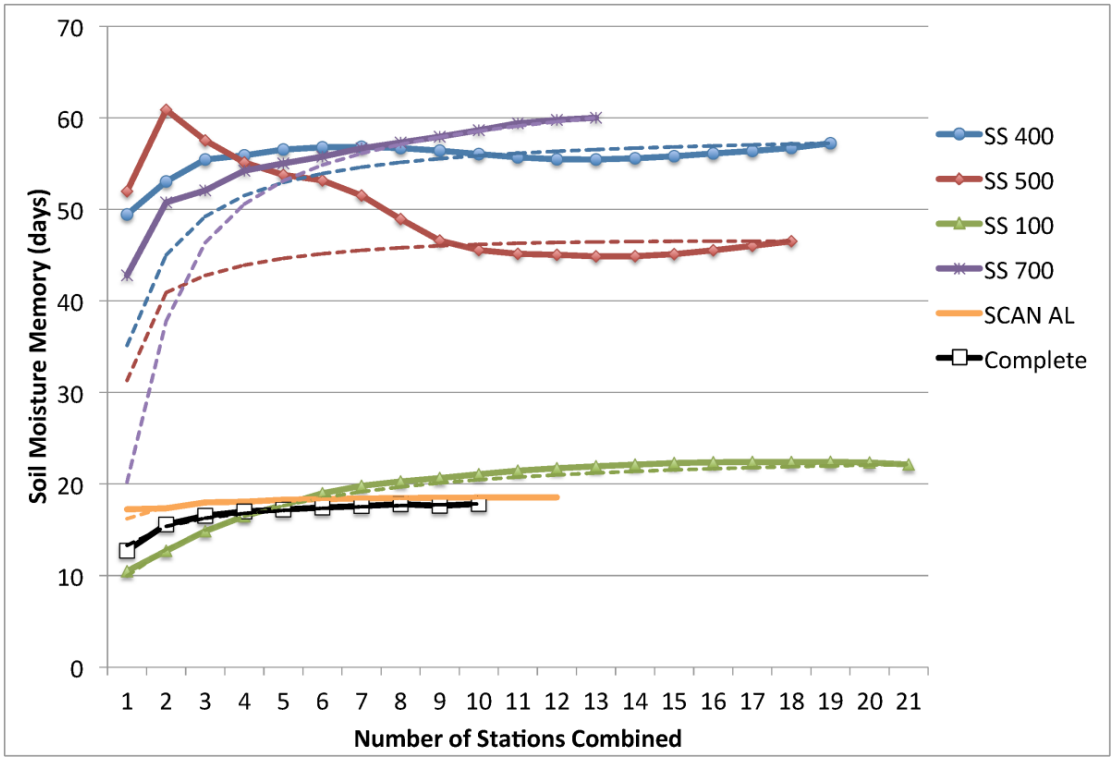

Figure 4: As in Fig 3 for soil moisture memory (see text for definition). Solid lines are the median value of memory among all combinations, and dotted lines show the harmonic mean across all combinations.

Network	*Network Mean*	GLDAS	CFSR	CFS	20CR	MERRAland	MERRA	GEOS5	ERA-I land	ERA-Interim	IFS	CLM	CCSM	Average Bias	*SD*
Ameriflux	**0.025**	-0.005	-0.006	-0.007	-0.013	-0.014	-0.015	-0.009	0.009	-0.013	0.005	0.016	0.003	**-0.004**	*0.010*
ARM	**0.016**	0.005	0.004	0.004	-0.002	-0.009	-0.011	-0.005	0.028	-0.003	0.024	0.047	0.015	**0.008**	*0.017*
AWDN	**0.010**	0.000	-0.002	-0.003	-0.014	-0.013	-0.015	-0.007	0.020	-0.008	0.015	0.034	0.009	**0.001**	*0.015*
CHILI	**0.026**	-0.012	-0.014	-0.013	-0.021	-0.022	-0.022	-0.016	0.036	0.007	0.031	0.004	-0.007	**-0.004**	*0.019*
COSMOS	**0.011**	0.000	-0.001	-0.003	-0.007	-0.007	-0.007	-0.004	0.017	-0.008	0.009	0.020	0.007	**0.001**	*0.009*
DEOS	**0.012**	0.000	0.009	0.002	0.005	-0.002	-0.005	-0.001	0.010	-0.005	0.004	0.027	0.020	**0.005**	*0.009*
ECONET	**0.021**	-0.013	-0.012	-0.018	-0.019	-0.021	-0.024	-0.017	-0.001	-0.018	-0.002	0.004	-0.002	**-0.012**	*0.009*
MAWN	**0.016**	-0.003	-0.003	-0.010	-0.010	-0.020	-0.021	-0.016	-0.007	-0.021	-0.005	0.020	0.008	**-0.007**	*0.012*
MAW-MO	**0.039**	-0.036	-0.033	-0.037	-0.041	-0.045	-0.048	-0.032	-0.015	-0.040	-0.013	-0.005	-0.020	**-0.031**	*0.013*
NOAAHMT	**0.013**	-0.001	-0.002	-0.001	-0.009	-0.002	-0.002	0.000	0.011	-0.009	-0.003	0.007	0.002	**-0.001**	*0.005*
OK-Meso	**0.013**	0.006	0.005	0.004	0.000	-0.007	-0.009	-0.001	0.026	-0.004	0.024	0.044	0.016	**0.009**	*0.015*
PBO-H2O	**0.012**	0.004	0.002	0.006	0.002	0.003	0.001	0.000	0.014	-0.003	0.006	0.015	0.007	**0.005**	*0.005*
SCAN	**0.015**	-0.007	-0.008	-0.010	-0.013	-0.014	-0.015	-0.013	0.005	-0.017	0.000	0.008	-0.003	**-0.007**	*0.008*
SDAWN	**0.011**	-0.007	-0.012	-0.010	-0.018	-0.022	-0.025	-0.021	0.009	-0.018	0.005	0.022	-0.002	**-0.008**	*0.014*
SNOTEL	**0.019**	-0.010	-0.009	-0.008	-0.012	-0.015	-0.019	-0.015	0.002	-0.017	-0.008	0.004	-0.006	**-0.009**	*0.007*
SOILSCAPE	**0.014**	0.003	0.001	0.009	0.000	0.003	0.005	0.008	0.012	-0.001	0.007	0.012	0.007	**0.006**	*0.004*
USCRN	**0.017**	-0.007	-0.008	-0.009	-0.012	-0.016	-0.017	-0.014	0.005	-0.017	0.000	0.013	0.000	**-0.007**	*0.009*
USDA-ARS	**0.011**	0.000	0.000	-0.003	-0.006	-0.007	-0.009	-0.008	0.018	-0.009	0.006	0.017	0.003	**0.000**	*0.009*
WTX-Meso	**0.027**	-0.014	-0.021	-0.023	-0.023	-0.032	-0.033	-0.033	0.006	-0.029	-0.007	0.012	-0.010	**-0.017**	*0.014*
Average	**0.017**	**-0.005**	**-0.006**	**-0.007**	**-0.011**	**-0.014**	**-0.015**	**-0.011**	0.011	**-0.012**	**0.005**	0.017	**0.002**	-0.004	
SD	*0.007*	*0.009*	*0.010*	*0.011*	*0.010*	*0.012*	*0.012*	*0.010*	*0.012*	*0.011*	*0.011*	*0.013*	*0.009*		

Figure 5: Average standard deviation of daily surface volumetric soil moisture (dimensionless) during JJA for stations in each network (first column after network names), and the average bias of standard deviation for each model across the station locations of each network. Averages and standard deviations of each network and model are shown at the right and bottom of the grid respectively. Color shading is to aid recognition of negative (blue) and positive (red) biases. Two networks that employ heat-dissipation sensors are shaded orange; two networks that use near-field remote sensing techniques are shaded green; all others use some form of dielectric probe.

Network	*Network Mean*	GLDAS	CFSR	CFS	20CR	MERRAland	MERRA	GEOS5	ERA-I land	ERA-Interim	IFS	CLM	CCSM	Average Bias	*SD*
Ameriflux	**7.6**	-0.7	-0.8	-0.5	14.1	11.0	5.0	7.4	7.7	-1.7	-0.3	-2.0	-5.6	**2.8**	*5.8*
ARM	**9.7**	-3.8	-2.6	-3.5	2.9	-2.5	-2.0	-2.3	8.3	-4.2	-4.1	-3.9	-4.3	**-1.8**	*3.6*
AWDN	**3.7**	2.8	5.5	2.9	10.1	9.8	9.5	10.7	13.7	2.0	1.7	2.5	-0.7	**5.9**	*4.4*
CHILI	**4.1**	5.4	7.4	4.9	14.7	9.9	9.6	12.9	19.4	0.4	0.3	4.6	1.0	**7.5**	*5.7*
COSMOS	**8.8**	-2.0	1.2	-1.9	5.3	7.1	6.6	9.4	8.4	-1.6	-2.0	-0.5	-5.1	**2.1**	*4.8*
DEOS	**15.9**	-7.3	-2.3	-6.1	3.0	-0.4	6.3	11.5	14.4	-8.1	-10.7	-8.9	-11.7	**-1.7**	*8.4*
ECONET	**7.2**	4.3	1.0	3.7	18.4	20.2	19.3	22.2	25.9	1.2	0.8	4.4	-3.7	**9.8**	*10.0*
MAWN	**8.7**	-2.5	0.7	-2.4	3.9	6.4	5.3	6.9	9.3	-1.9	-2.2	-0.9	-4.9	**1.5**	*4.4*
MAW-MO	**7.3**	-1.1	3.4	1.0	9.0	9.6	6.0	9.9	11.2	-2.2	-3.1	-0.7	-4.8	**3.2**	*5.5*
NOAAHMT	**7.9**	-1.1	-0.6	0.1	14.2	10.9	4.8	7.6	7.4	-1.7	-0.4	-2.5	-5.9	**2.7**	*5.9*
OK-Meso	**9.0**	-3.6	-2.6	-2.1	7.1	-1.2	-1.8	0.7	4.6	-4.8	-4.1	-4.1	-6.7	**-1.6**	*3.8*
PBO-H2O	**8.9**	-1.2	0.1	-3.8	-1.4	16.3	6.2	20.1	4.0	-3.5	-1.4	-0.7	-6.1	**2.4**	*7.8*
SCAN	**17.6**	-7.3	-0.5	-6.5	6.7	5.1	6.0	4.3	11.7	-8.8	-11.0	-9.7	-13.6	**-1.9**	*8.1*
SDAWN	**7.8**	-1.6	0.7	-2.2	5.8	7.9	6.2	6.8	10.4	-1.3	-1.4	-1.0	-4.9	**2.1**	*4.8*
SNOTEL	**14.0**	-9.3	-3.3	-7.9	0.2	23.3	30.9	18.3	0.1	-7.9	-8.3	-5.1	-12.4	**1.5**	*13.8*
SOILSCAPE	**9.6**	-3.9	-2.0	-5.0	0.0	7.5	7.3	8.7	5.9	-1.5	-1.7	-3.0	-5.1	**0.6**	*5.0*
USCRN	**14.3**	-8.2	-6.8	-6.6	-2.3	-2.8	-5.4	-5.1	6.9	-7.5	-7.6	-7.9	-11.0	**-5.4**	*4.3*
USDA-ARS	**6.7**	3.8	5.1	1.7	14.6	28.5	31.0	19.4	9.0	0.8	0.7	1.6	-4.8	**9.3**	*11.1*
WTX-Meso	**3.7**	0.7	7.4	1.2	5.5	30.2	32.2	13.1	0.0	0.0	0.0	12.0	-0.7	**8.5**	*11.1*
Average	**7.9**	**-1.9**	**0.6**	**-1.7**	**6.9**	**10.4**	**9.6**	**9.6**	**9.4**	**-2.7**	**-2.9**	**-1.4**	**-5.8**	**2.5**	
SD	*3.8*	*4.1*	*3.7*	*3.6*	*5.9*	*9.4*	*10.6*	*7.1*	*6.0*	*3.2*	*3.8*	*5.0*	*3.8*		

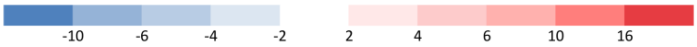

Figure 6: As in Fig 5 for surface soil moisture memory (units of days).

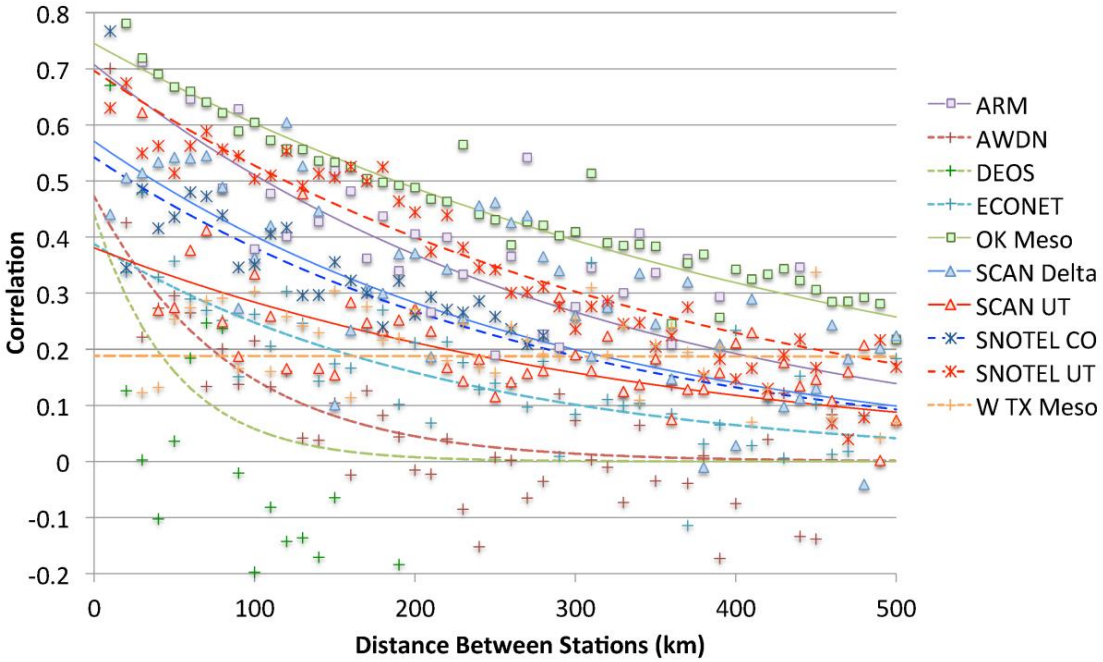

Figure 7: Spatial autocorrelation of daily surface layer soil moisture during JJA as a function of separation between stations for several networks, binned in intervals of 10km. Lines indicate linear best fits through the binned log of correlation (negative values omitted).

Network	Network Distance	GLDAS	CFSR	CFS	20CR	MERRAland	MERRA	GEOS5	Era-I land	ERA-Interim	IFS	CLM4	CCSM	Harmonic Mean	Mean	SD
ARM	307	164	546	536	1007	607	703	533	608	538	1054	751	858	659	529	*230*
AWDN	85	209	489	444	579	509	691	320	603	558	792	755	617	547	480	*162*
DEOS	49	326	512	421	670	635	737	387	635	607	590	959	547	585	541	*162*
ECONET	224	329	580	445	557	719	701	329	646	628	625	1004	530	591	538	*175*
OK-Meso	471	150	506	553	953	675	679	480	751	656	867	796	832	658	522	*207*
SCAN-Delta	285	468	734	501	637	601	660	303	928	813	860	1034	671	684	616	*197*
SCAN-UT	341	791	264	476	656	410	364	272	487	582	750	660	521	519	460	*167*
SNOTEL-CO	283	264	289	398	629	356	239	224	518	557	639	570	344	419	366	*149*
SNOTEL-UT	360	791	264	476	656	410	364	272	487	582	750	660	521	519	460	*167*
WTX-Meso	61518	163	493	519	845	531	689	476	819	686	748	725	691	615	509	*182*
Average		365	468	477	719	545	583	360	648	621	767	791	613	580	502	
Corr (no WTX)		0.22	-0.23	0.67	0.52	-0.12	-0.37	0.15	0.04	0.16	0.45	-0.35	0.31	0.14	-0.11	
Probability (1-tailed)		57%	#N/A	5%	15%	#N/A	#N/A	70%	91%	68%	23%	#N/A	42%	72%	#N/A	

Figure 8: Average spatial decorrelation distance (km) where zero-lag temporal correlation of time series of daily soil moisture drops to $1/e$ for surface volumetric soil moisture during JJA for stations in each network (first column after network names), and models. Colors indicate percentage deviation of each model (averaged over domain of network) relative to *in situ* network estimates with estimated random measurement error removed. *SD* is standard deviation among models.